

Gene Coexpression Networks in Human Brain Identify Epigenetic Modifications in Alcohol Dependence

Igor Ponomarev,^{1*} Shi Wang,^{1,2*} Lingling Zhang,^{1,2} R. Adron Harris,¹ and R. Dayne Mayfield¹

¹Waggoner Center for Alcohol and Addiction Research and the College of Pharmacy, University of Texas at Austin, Austin, Texas 78712, and ²Key Laboratory of Marine Genetics and Breeding, College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China

Alcohol abuse causes widespread changes in gene expression in human brain, some of which contribute to alcohol dependence. Previous microarray studies identified individual genes as candidates for alcohol phenotypes, but efforts to generate an integrated view of molecular and cellular changes underlying alcohol addiction are lacking. Here, we applied a novel systems approach to transcriptome profiling in postmortem human brains and generated a systemic view of brain alterations associated with alcohol abuse. We identified critical cellular components and previously unrecognized epigenetic determinants of gene coexpression relationships and discovered novel markers of chromatin modifications in alcoholic brain. Higher expression levels of endogenous retroviruses and genes with high GC content in alcoholics were associated with DNA hypomethylation and increased histone H3K4 trimethylation, suggesting a critical role of epigenetic mechanisms in alcohol addiction. Analysis of cell-type-specific transcriptomes revealed remarkable consistency between molecular profiles and cellular abnormalities in alcoholic brain. Based on evidence from this study and others, we generated a systems hypothesis for the central role of chromatin modifications in alcohol dependence that integrates epigenetic regulation of gene expression with pathophysiological and neuroadaptive changes in alcoholic brain. Our results offer implications for epigenetic therapeutics in alcohol and drug addiction.

Introduction

Brain gene expression is a critical determinant of brain function, including brain disease. Since the introduction of microarray technology, numerous studies have used transcriptome profiling to investigate the mechanisms underlying brain plasticity and brain pathology (Geschwind and Konopka, 2009). Alcohol and other drugs of abuse cause widespread changes in gene expression in human brain (Mayfield et al., 2008; Zhou et al., 2011), some of which contribute to the development and maintenance of drug dependence. Microarray studies in humans and animal models identified individual genes as mechanistic candidates for addiction phenotypes (Mayfield et al., 2008; Maze et al., 2011; Zhou et al., 2011), but an integrated view of molecular and cellular changes underlying alcohol and

drug addiction is lacking. Most genomic studies to date focused on individual genes with the highest statistical significance, limiting their discoveries to a handful of candidates. In some cases, this strategy resulted in mechanistic discoveries, but the bias toward most significantly regulated genes often lacks the functional foundation and contextual information for generating scientifically sound hypotheses.

Recent developments in statistical genomics and gene annotations provide a foundation for a shift from gene-centric to network- or module-centric systems approaches in data analysis. This shift is warranted by three key findings from recent literature on brain transcriptomes: (1) individual populations of neurons and glia are characterized by unique transcriptional signatures that reflect current functional state of these cells, (2) transcriptomes from complex tissues, such as whole brain, are organized into networks or modules of coexpressed (correlated) genes, and (3) these modules of coexpressed genes often reflect functional and structural organization of brain tissue and can be explained by known biological categories, such as cell type, cell organelles, and synaptic functions (Sugino et al., 2006; Lein et al., 2007; Cahoy et al., 2008; Doyle et al., 2008; Miller et al., 2008; Oldham et al., 2008; Ponomarev et al., 2010; Day and Sweatt, 2012). These discoveries advanced our understanding of organizational principles of brain transcriptomes and provided a biologically relevant context for interpreting differential expression of individual genes associated with CNS plasticity and pathology, offering critical insight into the mechanisms of Alzheimer's disease (Miller et al., 2010), schizophrenia (Torkamani et al., 2010), and posttraumatic stress disorder (Ponomarev et al., 2010).

Received June 20, 2011; revised Nov. 22, 2011; accepted Dec. 14, 2011.

Author contributions: I.P., S.W., and R.D.M. designed research; S.W. performed research; I.P., S.W., L.Z., and R.D.M. analyzed data; I.P., S.W., R.A.H., and R.D.M. wrote the paper.

This work was supported by National Institutes of Health/National Institute on Alcohol Abuse and Alcoholism Grants AA012404 (R.A.H.), Integrative Neuroscience Initiative on Alcoholism Grants AA013518 (R.A.H.), AA016648 (R.D.M.), AA013517 Pilot Projects (R.D.M. and I.P.), AA013476 (subcontract to I.P.), and K Award AA017234 (I.P.). The authors are grateful to the Australian Brain Donor Programs New South Wales Tissue Resource Centre for providing alcoholic and control brain tissue for analysis. The centre is supported by the University of Sydney, National Health and Medical Research Council of Australia, Schizophrenia Research Institute, and the National Institute of Alcohol Abuse and Alcoholism. We thank Dr. Vishy Iyer for critical reading of this manuscript and Brianne Patton and Elizabeth Osterdorff-Kahanek for technical assistance.

*I.P. and S.W. contributed equally to this work.

The authors declare no conflict of interest.

Correspondence should be addressed to either Igor Ponomarev or R. Dayne Mayfield, The University of Texas at Austin, Waggoner Center for Alcohol and Addiction Research, 1 University Station, A4800, Austin, TX 78712. E-mail: piatut@mail.utexas.edu, dayne.mayfield@mail.utexas.edu.

DOI:10.1523/JNEUROSCI.3136-11.2012

Copyright © 2012 the authors 0270-6474/12/321884-14\$15.00/0

To generate an integrated view of the brain transcriptome in human alcoholism, we profiled gene expression levels in post-mortem brains of human alcoholics and matched control cases and used a systems approach to data analysis that combined differential gene expression, gene coexpression networks, cell-type-specific transcriptomes, and a wide range of gene annotations. As a result, we identified critical epigenetic components in gene coexpression and proposed a central role for epigenetic regulation in alcohol-induced changes in global gene expression. Our approach allowed us to generate a unique systems hypothesis of brain changes in human alcoholism that integrated the epigenetic regulation of gene expression with previously reported cellular abnormalities. Our results offer implications for relevant treatment strategies and may serve as a prototype for analysis of the wealth of existing and emerging microarray data.

Materials and Methods

Case selection and postmortem tissue collection

Autopsy brain samples were obtained from the New South Wales Tissue Resource Centre at the University of Sydney. The Centre is funded in part by the National Institute on Alcohol Abuse and Alcoholism to provide brain tissue for alcoholism research. Fresh-frozen sections of tissue from the central (CNA) and basolateral nucleus (BLA) of amygdala, as well as the superior frontal cortex (CTX), were obtained from 32 cases (17 alcoholics and 15 matched controls; 30 males and 2 females). These regions are important substrates in the reward circuitry that is involved in the development of alcohol dependence and alcoholism (Koob and Volkow, 2010). Cases were matched as closely as possible by age, gender, postmortem interval (PMI), and brain pH. Diagnoses were confirmed by physician interviews, review of hospital medical records, questionnaires to next-of-kin, and pathology, radiology, and neuropsychology reports. Cases were also chosen on the basis that agonal hypoxia did not appear to have differed significantly from the study group. Moreover, none of the brains showed evidence of hypoxic encephalopathy, further suggesting that agonal hypoxia was minimal. We did not accept cases that suffered prolonged agonal states. Cases with a history of polydrug abuse were excluded. Cases were matched for smoking history. In addition, cases with concomitant diseases, such as cirrhosis of the liver, Korsakoff psychosis, or Wernicke or hepatic encephalopathies, were excluded. The concentration and quality of all RNA samples was determined, and degraded (RNA integrity number, RIN <4) and/or contaminated RNA samples were excluded from the analysis. The Diagnostic and Statistical Manual of Mental Disorders-IV diagnosis assigned to each case was based on a detailed and standardized clinical assessment summary created by specially trained staff with a background in psychiatry and/or psychology.

Microarray profiling and qRT-PCR validation

Total RNA was extracted and used to generate biotin-labeled cRNA using the Illumina TotalPrep RNA Amplification kit (Ambion). Biotin-labeled cRNA was then hybridized to Illumina HumanHT-12 whole-genome expression bead chips (Illumina). The quality of the Illumina bead summary data was assessed using the Bioconductor package Lumi. Data preprocessing included variance stabilization and quantile normalization. To eliminate potentially confounding effects of RNA quality on gene expression, we calculated residuals from the regression analysis of RIN values on gene expression and used them for statistical analysis and weighted gene coexpression network analysis (WGCNA) network construction. We next removed outlier values for each gene within a group using Grubbs' test ($p < 0.05$). Statistical analysis comparing alcoholic and control groups was performed using the Bioconductor package Limma to perform a Bayesian two-tailed t test. A false discovery rate (FDR) for each list of significantly regulated genes with nominal p values <0.05 was estimated using the method of Benjamini and Hochberg (1995). Our systems approach to prioritizing individual genes is based on integration of nominal statistical significance, gene network information, and functional relevance. Therefore, to avoid omitting true positives, all genes with nominal p values <0.05 were considered. After

initial data processing, microarray data from three brain regions of 15 controls and 17 alcoholics were used for network construction.

qRT-PCR validation. Several genes were selected for technical validation using qRT-PCR. qRT-PCR was conducted using amplified RNA from the same samples used for microarray experiments. Expression of *GIPC1* and *DNMT1* was examined in BLA and CTX, and expression of *MBD3*, *MLL4*, and *SETD1* was examined in BLA. All real-time TaqMan assays are pre-designed by Applied Biosystems and labeled with FAM as a reporter and a non-fluorescent quencher. Detailed TaqMan protocols are available on the manufacturer's website (<http://www.appliedbiosystems.com/>). *GUSB* was used as endogenous control, and the $\Delta\Delta C_t$ method was used to analyze data. Outlier values for each gene within a group were then removed based on the Grubbs' test ($p < 0.05$), and alcoholic and control groups were compared using a one-tailed t test.

WGCNA

Network construction. The general framework of WGCNA has been described in detail previously (Zhang and Horvath, 2005). We constructed a signed network for each brain region. Briefly, Pearson's correlations were calculated for all pairs of genes, and then a signed similarity (S_{ij}) parameter was derived: $S_{ij} = (1 + \text{cor}(x_i, x_j))/2$, where gene expression profiles x_i and x_j consist of the expression of genes i and j across multiple microarray samples. In the signed network, the similarity between genes reflects the sign of the correlation of their expression profiles. The signed similarity (S_{ij}) was then raised to power β to represent the connection strength (a_{ij}): $a_{ij} = S_{ij}^\beta$. This step aims to emphasize strong correlations and reduce the emphasis of weak correlations on an exponential scale. Here we chose a power of $\beta = 12$ for all brain regions so that the resulting networks exhibited approximate scale-free topology. Next, all genes were hierarchically clustered based on a dissimilarity measure of topological overlap which measures inter-connectedness for a pair of genes (Zhang and Horvath, 2005). The resulting gene dendrogram was used for module detection with the dynamic tree cut method (minimum module size, 100; cutting height, 0.99; and deepSplit = True). Gene modules corresponding to the branches cut off of the gene tree were labeled in unique colors. Unassigned genes were labeled in gray.

Functional annotation of genes and overrepresentation analysis of modules. We used several complementary approaches to characterize gene modules. For DAVID annotation, for each module, all genes in the module were submitted to the Database for Annotation, Visualization and Integrated Discovery (DAVID; <http://david.abcc.ncifcrf.gov/>). For cell-type annotation, gene sets known to be preferentially expressed in mouse oligodendrocytes, astrocytes, and neurons were obtained from Tables S4–S6 of Cahoy et al. (2008). We restricted our analysis to genes with at least fourfold enrichment in a given cell type. Another set of genes preferentially expressed in microglia was obtained from Table ST3 of Oldham et al. (2008). For each brain region, cell-type enrichment analysis was performed for all modules using the hypergeometric test with gene symbols as unique identifiers. For TE annotation, to identify microarray probes targeting potential transposable element (TE) regions, we first obtained genomic coordinates of all probes of Illumina from the probe annotation file HumanHT-12_V3_0_R2_11283641_A (<http://www.switchto.com/annotationfiles.ilmn>) and those of TEs from the University of California, Santa Cruz (UCSC) Genome Bioinformatics website (<http://genome.ucsc.edu>; genome assembly, NCBI36/hg18; track, RepeatMasker). Then we searched for probes targeting to TE regions by requiring that such a probe should fall completely into an annotated TE region. For each brain region, TE enrichment analysis was also performed for all modules and all major TE classes [i.e., DNA, long-term repeat (LTR), long interspersed nuclear element (LINE), short interspersed nuclear elements (SINE)] using the hypergeometric test. For GC content analysis, GC content (GC%) values for known genes were obtained from Ensembl (<http://www.ensembl.org>) using BioMart data management system. Gene GC content includes average GC% of the whole gene including exons, introns, and untranslated regions. GC% for each 50-mer Illumina probe was calculated based on the proportion of G and C nucleotides. Effects of gene GC content on gene coexpression was examined with a one-way ANOVA for average gene GC% calculated for each of 72 coexpression modules across all three brain regions.

Identification of alcohol-responsive modules. To identify potential alcohol-responsive modules, we used an effect size-based approach and determined the direction and magnitude of alcohol-induced changes for each coexpression module. We used the *t* statistics from the comparison of alcohol and control groups and calculated average *t* values for all genes within each module. We defined a “significant” alcohol-responsive module by requiring average $|t| > 1.8$ and a “suggestive” alcohol-responsive module by requiring average $|t|$ being between 0.9 and 1.8. In addition, we used a hypergeometric test to determine the degree of overrepresentation of differentially expressed genes in each module. All significant modules and a majority of suggestive modules were also significantly overrepresented with genes differentially regulated between the groups.

Meta-network analysis across brain regions. Module comparison between brain region networks was done following the approach described by Oldham et al. (2008). Briefly, for each pair of networks, the overlap between all possible pairs of modules was calculated, and the significance of module overlap was assessed using a one-sided hypergeometric test. The software Cytoscape (<http://www.cytoscape.org/>) was used to visualize the comparisons and create a meta-network of highly overlapping modules.

Control for confounding variables. In addition to correction for the effects of RIN by regression as described above, we determined whether confounding variables, such as age, gender, smoking history, PMI, or brain pH, contributed to differential gene expression between alcoholic and control cases. We obtained the first eigengene for each module (generated by WGCNA) that could explain the largest portion of variance in gene expression and then correlated these eigengenes with the confounding variables using Pearson’s correlation. The resulting *p* values were corrected for multiple comparisons with a Bonferroni’s procedure. None of the modules were significantly correlated with age, gender, or smoking history. Two modules were correlated with PMI (bla9, bla14) and three with brain pH (ctx4, ctx6, ctx16), and genes from these modules were excluded from consideration as alcohol-related candidate genes for future studies.

Chromatin assays

Quantification of global H3K4 histone methylation. Total histone extracts were prepared from CTX tissues of six alcoholics and six controls using the EpiQuik Total Histone Extraction kit (Epigentek) according to the instructions of the manufacturer. For global histone methylation quantification, monomethylations, dimethylations, and trimethylations of H3K4 were measured using the EpiQuik Global Pan-Methyl Histone Quantification kits (colorimetric assay) (Epigentek). For each sample, 1.5 μ g of histone extracts was used in each assay.

DNA methylation analysis of LTR retrotransposons. Genomic DNA was extracted from CTX tissues of six alcoholics and six controls using the DNeasy Blood and Tissue kit (Qiagen). In a total volume of 80 μ l, 80 ng of genomic DNA was digested with 30 U of McrBC (New England Biolabs) at 37°C for 3 h. The same reaction was set up for the mock digestion without adding of McrBC. Then digested and undigested DNA was subjected to real-time PCR quantification using primers for specific LTRs: MLT2A1 (forward, 5′-GAGAGGCAGACCCACCTTA-3′ and reverse, 5′-CACGATACAAGGTCCACAA-3′), LTR8 (forward, 5′-CAAGC-TGTCCTTGTTTCATTCCT-3′ and reverse, 5′-CTGCTTTGGGAAAGG-GCTGTT-3′), and THE1B (forward, 5′-TCATCTTGAATTGTAGC-TCCCAT-3′ and reverse, 5′-TCCCCTTTATAAACCATCAGAT-3′). Primers were designed based on the consensus LTR sequences retrieved from the RepBase 14.04. Real-time PCR amplification was set up in triplicate in a 20 μ l volume composed of 5 ng of digested or undigested DNA, 0.2 μ M of each primer, 1 \times Power SYBR Green PCR Master Mix (Applied Biosystems) in the 7900HT Fast Real-Time PCR System (Applied Biosystems). All cycling began with an initial denaturation at 95°C for 10 min, followed by 40 cycles of 95°C for 15 s, and 60°C for 1 min. The “mock-digestion” control was used to normalize the McrBC–qPCR data. Fold change was calculated by dividing the normalized data by the mean of the control group.

Chromatin immunoprecipitation assay. Chromatin immunoprecipitation (ChIP) assays were performed for five alcoholics and five controls using the EpiQuik Tissue Methyl-Histone H3K4 ChIP kit (Epigentek).

The ChIP-grade antibody, anti-trimethyl-H3K4 was purchased from the same company. The normal IgG served as a negative control in the ChIP assay. The ChIP protocol of the manufacturer was followed with minor changes. Briefly, for each sample, ~ 10 mg of CTX tissue was smashed and cross-linked in the presence of 1% formaldehyde for 15 min at room temperature. The cross-linking was then stopped by adding 1/10 volume of 1.25 M glycine solution. After being washed by 1 ml of ice-cold 1 \times PBS, tissue was homogenized using a motorized pestle (VWR) and pelleted through centrifugation at 10,000 rpm for 5 min at 4°C. Tissue pellet was then resuspended in a lysis buffer containing protease inhibitor cocktail. Chromatin was sheared by sonication on ice for 10 min in total (20 s ON, 40 s OFF) at level 2 using the Fisher Model 60 Sonic Dismembrator (Thermo Fisher Scientific). The length of sheared DNA was checked by agarose gel electrophoresis, which is usually between 200 and 1000 bp. Next, following the protocol of the manufacturer, 100 μ l of sheared chromatin was used for immunoprecipitation (IP) with anti-trimethyl-H3K4. For input-DNA preparation, 2 μ l of proteinase K was first added into 100 μ l of sheared chromatin and incubated at 65°C for 15 min. Then 4 μ l of 5 M NaCl was added into the input-DNA solution and incubated at 65°C for 2 h. After reversal of cross-linked DNA, IP and input DNA were purified through fast-spin columns. IP and input DNA were subjected to real-time quantitative PCR using primers specific to the promoter regions of one of six genes: *GIPCI* (forward, 5′-CCCCAGAGATTGAATGCATCTT-3′ and reverse, 5′-GATTTCGAACCTCCGACGTCCA-3′), *BCL2L1* (forward, 5′-TGAACCCCATTTGAGAAGTCCCT-3′ and reverse, 5′-ACTGGGAGCCAGGAGTACTCT-3′), *UBE1* (forward, 5′-CTTGACAGCCTGGCTGCAACA-3′ and reverse, 5′-TGCATAAAGTTCCTACTCGGT-3′), *ARHGDI1* (forward, 5′-CCTCACACTGCCCCAGAGGAT-3′ and reverse, 5′-GCGCACTTCTGAGCAGGAGT-3′), *CLPTM1* (forward, 5′-GGAAACAAACGGCTGGGAGA-3′ and reverse, 5′-CGCGAGATTTACGCTTTCTTA-3′), and *CALCOCO1* (forward, 5′-TGCCTGCGAGCCTTCTGGGAT-3′ and reverse, 5′-CAACAAAAACAGCACTCCGACT-3′). Real-time PCR amplification was set up in triplicate in a 20 μ l volume composed of 0.5 μ l of IP or input DNA, 0.2 μ M of each primer, and 1 \times Power SYBR Green PCR Master Mix (Applied Biosystems) in the 7900HT Fast Real-Time PCR System (Applied Biosystems). All cycling began with an initial denaturation at 95°C for 10 min, followed by 40 cycles of 95°C for 15 s, and 60°C for 1 min. PCR specificity was checked by melting curve analysis. The “input” control was used to normalize the ChIP-qPCR data. Fold change was calculated by dividing the normalized data by the mean of the control group.

Results

Alcohol abuse is associated with widespread changes in brain gene expression

To define alterations in the brain transcriptome produced by chronic alcohol abuse, whole-transcriptome gene expression profiling was conducted for three brain regions (BLA, CNA, CTX) from 17 alcoholics and 15 matched control cases. History of alcohol abuse was associated with global changes in gene expression in all three brain regions. “Global” here refers to the fact that numbers of transcripts differentially expressed at a nominal $p < 0.05$ in different brain regions [3589 for BLA (FDR < 20%), 2656 for CNA (FDR < 28%), and 2716 for CTX (FDR < 28%)] were statistically greater than those expected by chance (all hypergeometric $p < 0.0001$). Overall, our results corroborate previous studies showing widespread changes in brain gene expression in alcoholics (Mayfield et al., 2008; Zhou et al., 2011). These studies identified many candidate genes that may play a role in alcoholism, but our goal was to extend this line of research beyond the gene-centric approach and to generate and validate easily testable hypotheses at a systems level.

Construction and validation of gene coexpression networks

Our next step was to construct gene coexpression networks to gain insights into functional organization of the brain transcriptome. “Coexpression” here refers to the implication that genes whose expression covaries (is correlated) across samples are coexpressed, i.e.,

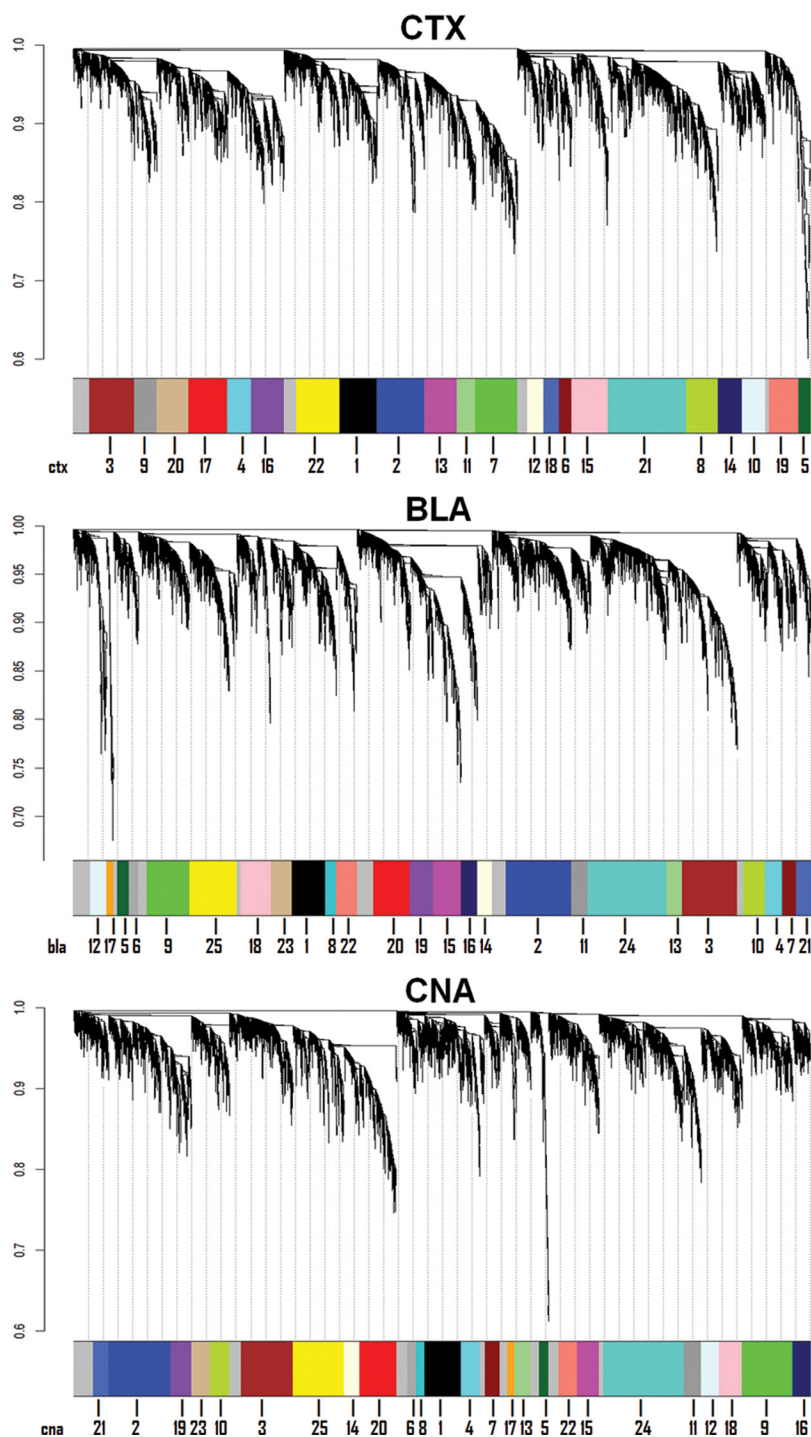


Figure 1. Network analysis of gene expression in three brain regions of human alcoholics and control cases identifies distinct modules of coexpressed genes. Shown are dendrograms produced by average linkage hierarchical clustering of transcripts (see Materials and Methods). Horizontal color bars represent different coexpression modules that are also numbered. Bar sizes correspond to the number of transcripts in each module.

regulated by similar mechanisms. Identification of gene coexpression patterns has been a fruitful approach to understanding mechanisms of transcriptional regulation in brain (Oldham et al., 2008). We used the WGCNA (Oldham et al., 2008) to construct a gene coexpression network for each of the three brain regions. This method has been described in detail previously (Oldham et al., 2008), and its utility as a systems tool has been validated by several research groups (Saris et al., 2009; Torkamani et al., 2010; Mulligan et al., 2011).

All reliably detected genes were included in the network construction, and data from both alcoholics and non-alcoholics were combined to detect coexpression patterns. In total, we identified 72 modules in three gene coexpression networks with 25 modules for BLA, 25 for CNA, and 22 for CTX (Fig. 1). The module size (i.e., total number of genes in a module) ranged from ~100 to ~1600. To evaluate biological significance of the modules, we used a wide range of gene annotations, including GO, KEGG, and major cell classes in brain (Cahoy et al., 2008; Oldham et al., 2008), and examined an overrepresentation (enrichment) of each biological category in a given module by comparing numbers of genes annotated with this biological category with chance (for details, see Materials and Methods). Most modules were highly overrepresented with at least one functional or structural category, thus validating biological relevance of gene coexpression relationships. Similar to previous reports (Oldham et al., 2008; Miller et al., 2010; Ponomarev et al., 2010), the most overrepresented biological categories in the present study included major cell classes, such as neurons, astrocytes, oligodendrocytes and microglia, and cell organelles, such as mitochondrion, nucleus, and ribosome (Table 1). This shows that cells and cellular compartments are main sources of gene coexpression and indicates that cell-type-specific transcriptional signatures can be obtained from complex brain tissue without isolating cellular populations. We next asked whether variation in chromatin states could contribute to gene coexpression.

Detecting epigenetic variation in gene coexpression

Understanding principles of modular organization in gene coexpression remains a challenge because many modules of highly coexpressed genes are not readily explained by cellular identity or any of the other commonly used annotated functions. This has not been a problem of annotation availability but rather a problem of annotation usage, because most gene expression studies do not explore expression patterns beyond traditionally used databases, such as GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes). One area in which additional effort is warranted is chromatin marks at individual gene locations. Changes in chromatin structure, often termed epigenetic changes, including DNA methylation and histone modifications, are critical variables affecting global gene expression. Therefore, it is reasonable to expect that coexpression of genes in some modules will be driven by chromatin changes.

Table 1. Biological categories critical for the organization of brain transcriptomes and gene coexpression relationships

Biological categories	<i>p</i> value
Cell classes	
Neuron	10^{-104}
Microglia	10^{-56}
Astrocyte	10^{-94}
Oligodendrocyte	10^{-54}
Cell Organelles	
Ribosome	10^{-57}
Mitochondrion	10^{-16}
Nucleus	10^{-9}
TEs	
LTR	10^{-17}
SINE	10^{-151}

Shown are categories highly overrepresented in at least one module in each brain region. Overrepresentation *p* values represent the highest statistical significance detected for a given biological category.

To explore the effects of chromatin state on gene coexpression relationships, we used two variables that are easily obtained from microarray data: expression of genomic repeats and gene GC content. Repeated sequences, most of which are represented by TEs of various classes, constitute a large fraction of most eukaryotic genomes. Transposons are homologous DNA fragments that are present in multiple copies in the genome and are capable of being reproduced and randomly inserted in the host genome (Slotkin and Martienssen, 2007). Transposons are normally silenced by epigenetic mechanisms, including DNA methylation, modifications of histone tails, and alterations in chromatin packing and condensation, but can be expressed when the epigenetic silencing is released (Slotkin and Martienssen, 2007). Therefore, expression of transposons may serve as a sensitive marker of changes in chromatin state.

We used the RepeatMasker program (see Materials and Methods) and found that 3992 Illumina microarray probes could be mapped to one of four classes of TEs, either DNA transposons or one of three types of RNA transposons (retrotransposons): LTR-containing endogenous retroviruses (ERVs), LINEs, or SINEs. Expression of 825 of these probes was statistically greater than the background noise in at least one brain region. We validated these results by manually checking the genomic location of ~15% of the probes using the UCSC genome browser. Most probes corresponding to TEs also mapped to untranslated regions or introns of known or predicted genes, whereas a relatively large fraction of LTR probes (~20%) also mapped to multiple intergenic regions (Fig. 2). Our overrepresentation analysis of coexpression modules identified several modules that showed significant enrichment with TEs in all brain regions. Coordinated expression of Illumina probes corresponding to LTRs and SINEs was of particular interest because several modules were highly statistically overrepresented with these TEs (Table 1). Many TEs have retained functional promoters, and the effects of TEs on expression of adjacent individual genes have been well documented (Waterland and Jirtle, 2003). Our overrepresentation results suggest, to our knowledge for the first time, that epigenetically controlled TEs can regulate multiple genes in a coordinated manner.

The second variable obtained from our microarray data was gene GC content, a measure of the nucleotide composition of the gene. Nucleotide composition of individual genes and their promoters plays a critical role in regulation of transcription; two examples include DNA methylation at CpG dinucleotides, a marker of transcriptional repression, and preferential binding of different transcription factors and other regulatory proteins to

either GC- or AT-rich motifs (Dekker, 2007). This notion is consistent with studies that reported robust correlations between genomic GC content and several epigenetic marks, including DNA methylation, some histone modifications, and chromatin condensation (Vinogradov, 2005; Koch et al., 2007). We next examined whether gene GC content contributed to gene coexpression. GC content (GC%) values for each gene were obtained from Ensembl, averages for each coexpression module were calculated, and one-way ANOVA was performed. Average GC% showed remarkable variability among modules, ranging from 40 to 56% (see Fig. 4A), and ANOVA resulted in a highly significant *p* value ($F_{(71, 34,145)} = 235$; $p < 10^{-500}$), indicating that gene GC content is a critical variable affecting gene coexpression and suggesting that genes with similar GC content are generally coregulated. Because both TEs and GC% are mechanistically associated with chromatin marks, our data point to previously unrecognized epigenetic sources in gene coexpression and suggest that coregulation of TEs and genes with similar GC content reflect individual variation in chromatin states and can be used as markers of epigenetic regulation of gene expression.

Coexpression patterns are highly conserved across brain regions

We next inquired whether observed coexpression patterns in three networks were conserved across brain regions. Module comparison between networks was accomplished by identifying overlapping genes and calculating statistical significance of the overlap between all possible pairs of modules. We found that all modules in a given brain region have genes significantly overlapping with at least one module from a different brain region, and the majority of modules were highly overlapping across all three brain regions (Fig. 3), suggesting conserved patterns of gene regulation in different brain regions. This finding was consistent with the study by Oldham et al. (2008) that showed similar coexpression preservation in three different regions. Similar to their study, we found that major brain cell classes, which include neurons, astrocytes, oligodendrocytes, and microglia, as well as cellular organelles, including mitochondria and ribosomes, are the most conserved biological categories with respect to gene coexpression. In addition, our analysis provides the first evidence that regulation of modules enriched with LTR and SINE TEs, as well as modules containing genes with high or low GC content, are conserved and cluster together. We hypothesize that conserved modules representing TEs and opposite ends of the GC% spectrum reflect fundamental epigenetic influences on gene coexpression relationships.

Gene coexpression networks provide insight into functional changes in alcoholic brain

By constructing gene coexpression networks and identifying biological sources of coexpression modules, we created a functional framework for interpretation of differential expression between alcoholics and controls at a systems level. To examine global effects of alcohol abuse on gene coexpression networks, we used an effect size-based approach and determined the direction and magnitude of alcohol-induced changes by calculating average *t* values for genes of each coexpression module (Fig. 3, shown in color). *t* tests were conducted for every transcript in each brain region to compare gene expression between alcoholics and control cases, and *t* values can be used as estimates of the effect size. This revealed three main findings: (1) chronic alcohol abuse differentially affected major cell types in brain: transcripts from neuronal modules were mainly downregulated in alcoholics,

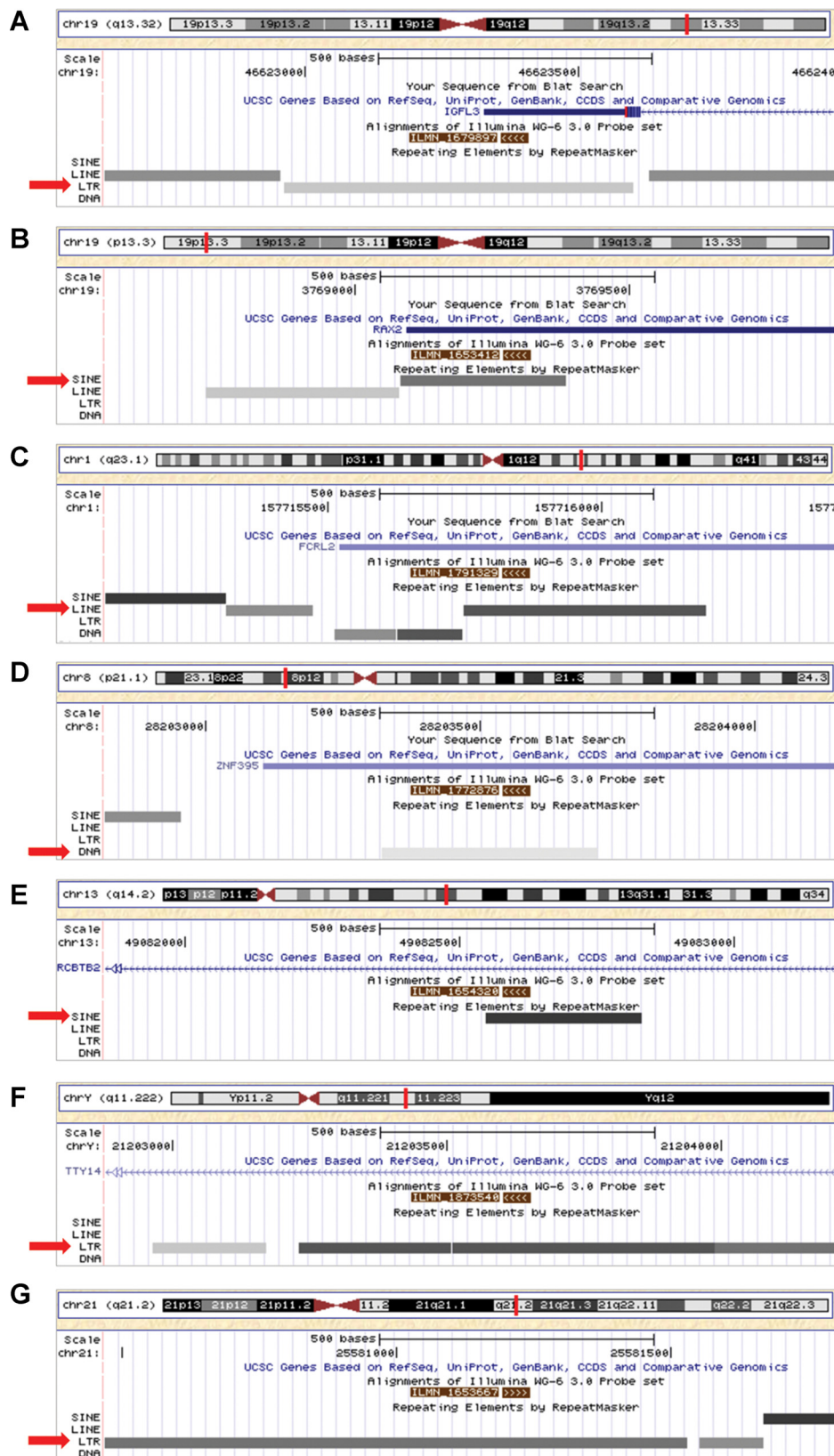


Figure 2. Illumina probes can detect expression of TEs in human genome. Shown are seven example panels from the UCSC Genome Browser (<http://genome.ucsc.edu>) showing genomic locations that are perfect matches for several Illumina probes representing TEs. Each panel shows whole chromosome on top, with the small red bar representing the location of the Illumina probe. The next track (in blue) shows the location of a known gene, which is followed by the location of the Illumina probe (in brown). Four tracks at the bottom represent the location of known or predicted TEs, including SINE, LINE, LTR, and DNA transposons. Red arrows point to TEs to which Illumina probes map. **A**, Probe mapping to an LTR transposon and a 3' UTR of a known gene (*IGFL3*). **B**, Probe mapping to a SINE transposon and a 3' UTR of *RAX2* gene. **C**, Probe mapping to a LINE transposon and a 3' UTR of *FCRL2* gene. **D**, Probe mapping to a DNA transposon and a 3' UTR of *ZNF395* gene. **E**, Probe mapping to a SINE transposon and an intron of *RC3TB2* gene. **F**, Probe mapping to an LTR transposon and an intron of *TTY14* gene. **G**, Probe mapping to an LTR transposon and an intergenic region.

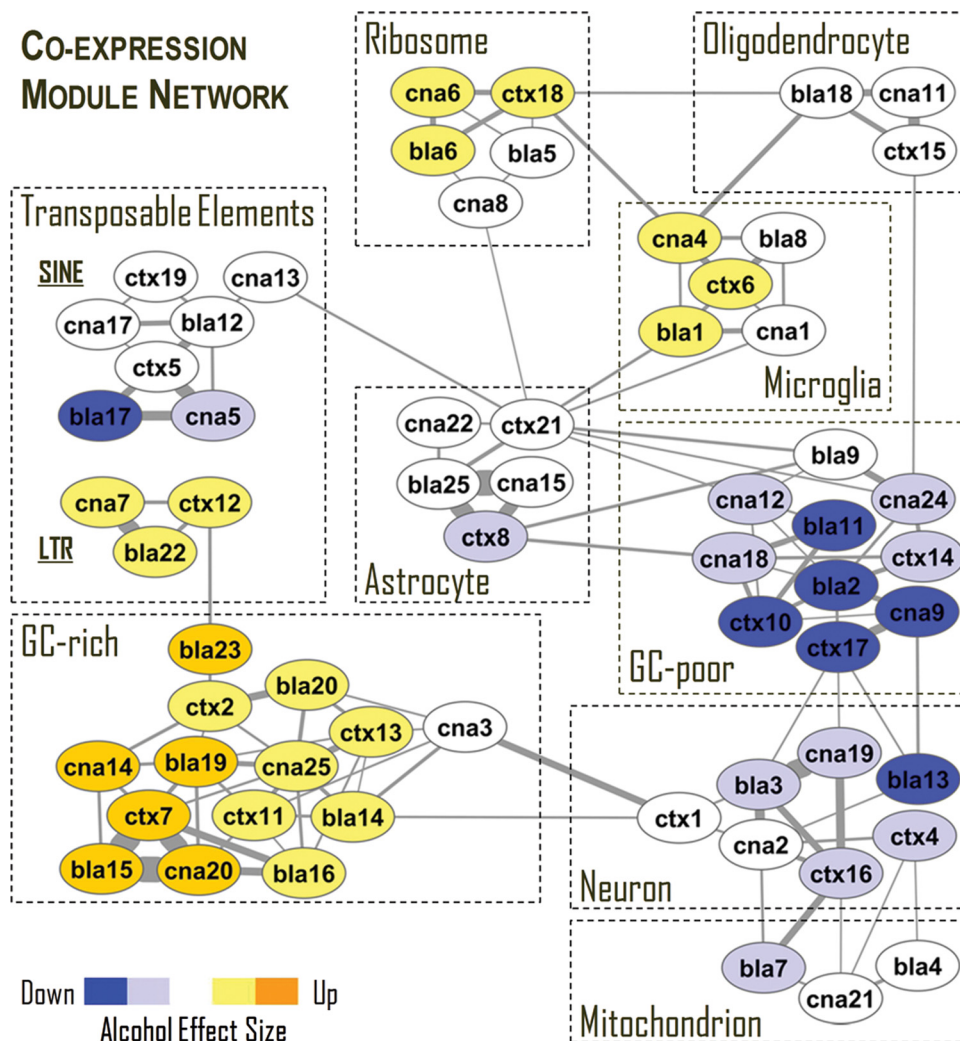


Figure 3. A meta-network of overlapping gene coexpression modules in human brain. Each node represents a module of coexpressed genes. Nodes are labeled with brain region and a module number. An edge between two nodes indicates a significant overlap of genes between two modules of different brain regions. Shown are only highly overlapping modules ($p < 10^{-20}$). An overlap of three and more modules from different brain regions (e.g., ribosomes) indicates a cluster of highly conserved coexpression modules (represented by rectangular boxes with a dashed borderline). All overlapping modules within a cluster are overrepresented with genes from a major biological category shown in Table 1, such as neuron or TEs. In addition, several modules were clustered based on average GC content of their genes; GC-rich and GC-poor modules formed separate clusters. Several GC-poor modules were also overrepresented with “nucleus” genes. Thickness of connecting edges is proportional to the significance of the overlap. Module colors represent the direction and magnitude of regulation in alcoholic brain based on average t values (see Materials and Methods) (yellow, upregulation; blue, downregulation in alcoholics; intense colors, $|t| > 1.8$; light colors, $|t| > 0.9$).

whereas several modules representing microglia were upregulated; (2) alcohol abuse resulted in upregulation of LTR retrotransposons; and (3) most genes from GC-rich modules were upregulated in alcohol abusers, whereas genes from GC-poor modules were mainly downregulated. This last finding was especially intriguing because it suggested that gene nucleotide composition determines, at least in part, whether genes will be regulated in response to strong environmental challenges, such as chronic alcohol abuse. We further investigated the relationship between gene GC content and regulation by chronic alcohol by calculating Pearson's correlation between average gene GC content and t values for the 72 coexpression modules (Fig. 4B). Remarkably, gene GC content accounted for $\sim 68\%$ ($r = 0.83$; $p < 10^{-18}$) of the differential gene expression between alcoholics and controls. This relationship was not an artifact of differential microarray probe hybridization, because neither average gene GC content nor average Illumina probe GC content correlated with average transcript expression values (Fig. 4C,D). Based on the rationale discussed above, the coordinated regulation of LTR ret-

rotransposons and genes with similar GC content suggests a critical role of chromatin modifications in the modulation of gene expression in the alcoholic brain. Based on the results of the integration of coexpression networks with differential gene expression, we further investigated the effects of alcohol abuse on cellular transcriptomes and chromatin modifications.

Effects of alcohol abuse on cell-type-specific transcriptomes

Neuronal and glial cells are the fundamental constituents of the CNS. Despite identical genomes, different cell types use distinct transcriptional programs that result in remarkable heterogeneity of cellular transcriptomes that are thought to reflect physiological properties and the functional state of individual cells (Sugino et al., 2006; Doyle et al., 2008). To investigate the effects of alcohol abuse on cell-type-specific gene expression, we again used the effect size-based approach and analyzed distributions of t values for genes that are primarily expressed in one of the four major cell classes: neurons, microglia, oligodendrocytes, and astrocytes (Fig. 5). Cell-type-specific genes were determined by literature

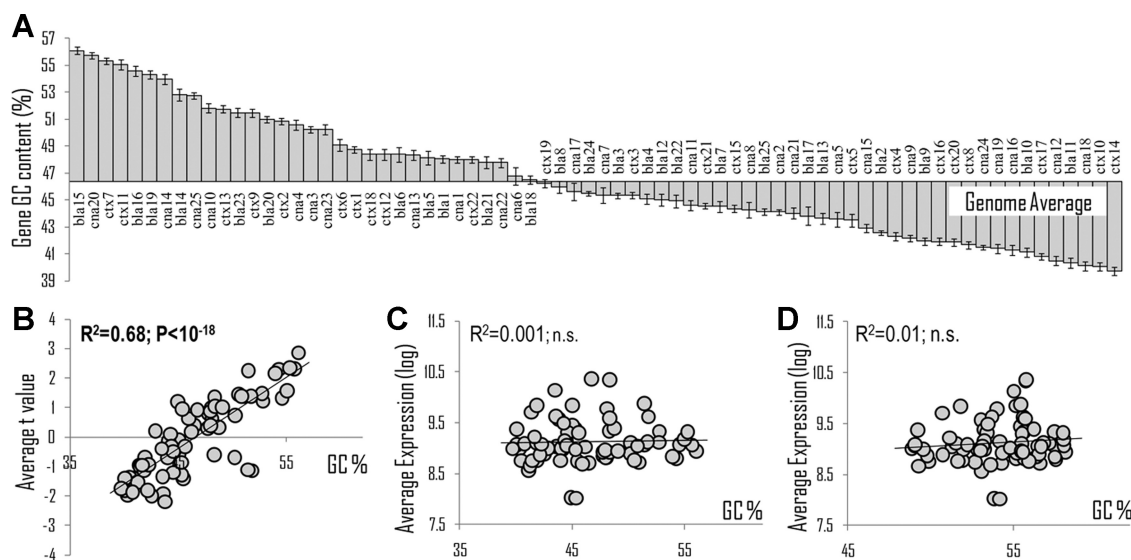


Figure 4. Effects of gene GC content on gene coexpression relationships and regulation by alcohol abuse. **A**, ANOVA of average gene GC content (GC%) for 72 coexpression modules (labeled with brain region and module number) reveals remarkable heterogeneity among modules ($F_{(71, 34, 145)} = 235; p < 10^{-500}$). **B**, Average gene GC% for 72 coexpression modules plotted versus average t values. A t value represents the magnitude and direction of alcohol effects ($t < 0$ indicates downregulation; $t > 0$ indicates upregulation by alcohol). Average gene GC% (**C**) and average GC% of Illumina probes (**D**) are plotted versus average gene expression values for the 72 coexpression modules (r = Pearson's correlation; n.s., not significant).

(Cahoy et al., 2008; Oldham et al., 2008; for details, see Materials and Methods). We hypothesized that the shape and position of the t distributions can reveal global effects of alcohol on individual cell types. In principle, an alcohol-induced change in expression of a particular gene reflects one of two distinct possibilities: an actual change in mRNA copy number or a change in the abundance of tissue or the number of cells in which this gene is preferentially expressed. For example, a significant shift of the t distribution mean or median compared with zero chance would most likely indicate a change in abundance or general activity of a cellular population, whereas deviation from normality as, for example, “bumps” on the distribution may indicate a coordinated expression of functionally relevant genes.

Our analysis revealed discrete effects of alcohol on different cell types in different brain regions (Fig. 5). Neuronal distributions in the amygdalar regions were significantly shifted to the left, whereas all microglial distributions were shifted to the right, suggesting a decrease in numbers of neurons and an increase in microglia. In addition, several molecular markers of activated microglia, such as *CCL2* and *TSPO*, were significantly upregulated in the amygdala (all $p < 0.02$), whereas neuronal markers, such as *SST*, *VIP*, and *GABRG2*, were generally downregulated. These results are consistent with alcohol literature showing general degeneration of neurons as well as activation and proliferation of microglia in alcoholic brain (Crews et al., 2011). This analysis also showed clear differences in regional sensitivity to alcohol because BLA was the most affected region, whereas CTX was the least affected.

Detailed analysis of the neuronal t distribution in cortex revealed a deviation from normality because several genes significantly upregulated in alcoholics contributed to a bump on the distribution (Fig. 5, arrow). Most of these genes were clustered in the GC-rich *ctx7* module (Fig. 3). A majority of the deviated genes were annotated as being involved in synaptic transmission, particularly at glutamatergic synapses; examples include dynamin (*DNM1*; $p = 0.007$), syntaxin (*STX1A*; $p = 0.04$), synapsin I (*SYN1*; $p = 0.05$), synaptophysin (*SYP*; $p = 0.005$), glutamate NMDA receptor NR1 subunit (*GRIN1*; $p = 0.008$), and vesicular

glutamate transporter 1 (*VGLUT1*, *SLC17A7*; $p = 0.001$). Two additional upregulated genes from the *ctx7* module with roles in glutamatergic neurotransmission are *GIPCI* ($p = 0.005$) and *MIB2* ($p = 0.0002$), which are involved in NMDA receptor trafficking and ubiquitination of the NMDA NR2B subunit, respectively (Yi et al., 2007; Jurd et al., 2008). Another striking discovery was that GC content of all of these genes was greater than average, suggesting that this played a role in coordinated upregulation of synaptic genes in alcohol abusers.

Activation of ERVs in alcoholic brain is associated with DNA hypomethylation

Detailed examination of three highly overlapping modules overrepresented with LTR transcripts (Fig. 3) revealed that the majority of the transcripts were upregulated in alcoholics, with many upregulated probes mapping to multiple intronic and intergenic genomic regions corresponding to LTR TEs. This pattern of expression is consistent with a genome-wide transcriptional activation of LTR retrotransposons in alcoholic brain. LTR-containing TEs represent a class of ERVs, most of which are nonfunctional remnants of ancient retroviral infections (Antony et al., 2004). However, many human ERVs (HERVs) have retained functional promoters and the potential to encode viral proteins, randomly insert their DNA in the genome, and modify the expression of adjacent genes (Morgan et al., 1999; Waterland and Jirtle, 2003). Because expression of ERVs can cause genomic instability and disease (Antony et al., 2004), eukaryotic hosts developed defense mechanisms against these genomic parasites. The LTR regions of ERVs are heavily methylated in somatic cells, which was proposed as a primary mechanism of their transcriptional repression (Waterland and Jirtle, 2003). Expression of ERVs correlates with subtle changes in DNA methylation status (Waterland and Jirtle, 2003), and ERV activity can be used as a sensitive marker of global DNA hypomethylation (Schulz et al., 2006; Balada et al., 2009).

Here, we tested the hypothesis that DNA in brains of alcoholics is less methylated, which results in transcriptional activation of HERVs. We used a qPCR-based method to measure DNA

methylation in CTX of alcoholic and control cases for three LTR families and observed a reduction of DNA methylation in the LTR region of these retrotransposons (Fig. 6A), suggesting that activation of ERVs in alcoholics was attributable, at least in part, to DNA hypomethylation. This finding was consistent with a 20–30% downregulation of the DNA methyltransferase *DNMT1* in all three brain regions of alcoholics (BLA, $p = 0.002$; CNA, $p = 0.05$; CTX, $p = 0.04$). *DNMT1* plays an important role in the establishment and regulation of tissue-specific patterns of methylated cytosine residues, and a reduction of *DNMT1* activity is often observed together with global DNA hypomethylation in several types of cancer and other pathological conditions (Hervouet et al., 2010). Alcohol-induced global DNA hypomethylation has been reported in liver (Lu et al., 2000), fetal tissue (Garro et al., 1991), and colon (Choi et al., 1999), and our study is the first to report it in human brain.

Upregulation of GC-rich genes in alcoholics is associated with increased H3K4 trimethylation

We next focused on modules containing GC-rich genes, many of which were upregulated in alcoholics (Fig. 3). Three of these genes significantly upregulated in all three brain regions (all $p < 0.05$) were the histone methyltransferases *MLL*, *MLL4*, and *SETD1A* specific for trimethylation of histone 3 at lysine 4 (H3K4me3), a chromatin mark of actively transcribed genes. Because of this upregulation and a strong positive correlation between genome GC content and the H3K4me3 mark (Koch et al., 2007), we hypothesized that upregulation of some genes from the GC-rich modules in alcoholics is associated with increased H3K4me3. First, we found that global trimethylation was increased in alcoholic brain (Fig. 6B). Next, we used ChIP-qPCR to test H3K4 trimethylation level at the promoter region of six hub genes from the *ctx7* module that were upregulated in alcoholics. Three of six genes (*GIPC1*, *BCL2L1*, and *UBE1*) showed significantly increased levels of H3K4 trimethylation in alcoholics (Fig. 6B), which was consistent with the upregulation of their transcripts, whereas H3K4me3 levels of the other three genes did not differ between the groups. These results suggest that the alcohol-induced upregulation of genes in the GC-rich modules may, at least in part, be explained by increased levels of H3K4me3 in their promoters.

Alcohol-induced upregulation of genes involved in transcription corepressor complexes

Upregulation of several functionally related genes point to another mechanism of epigenetic control. Methyl-CpG-binding protein *MBD3* and chromodomain helicase *CHD4* were significantly upregulated in alcoholics (*MBD3* in all brain regions; *CHD4* in BLA; all

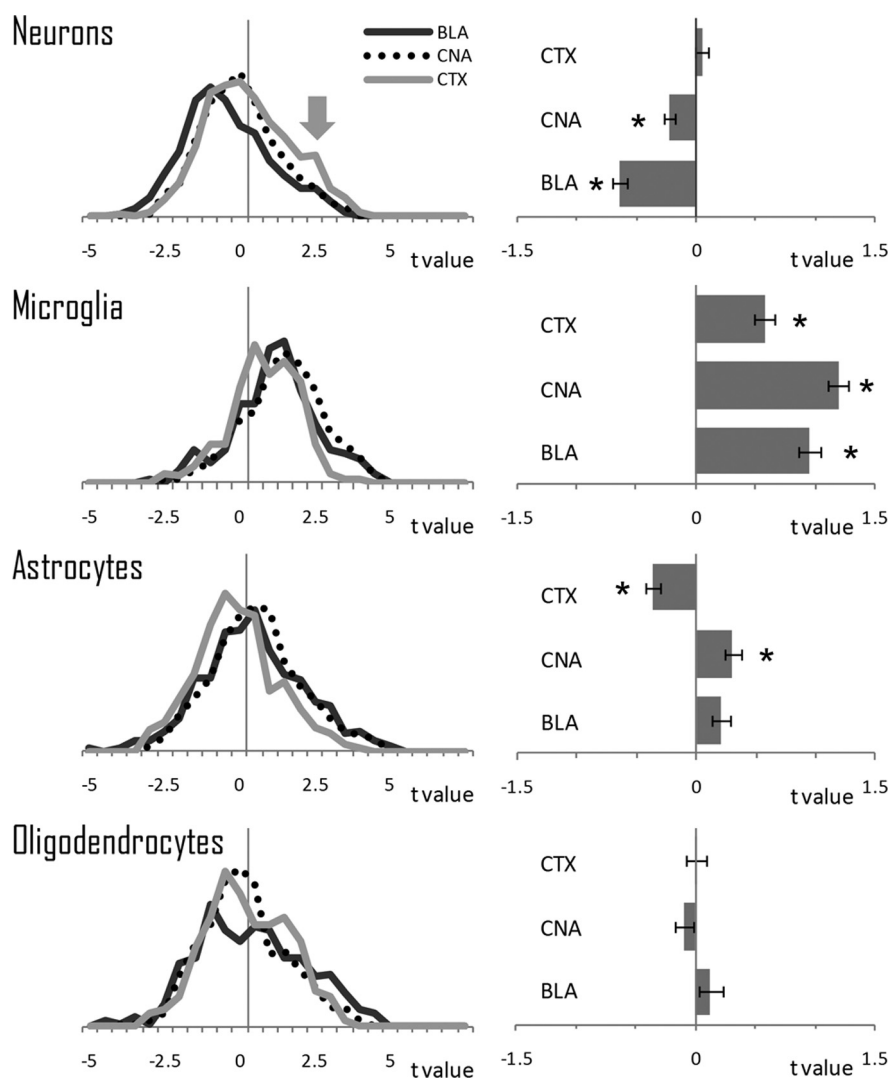


Figure 5. Global effects of alcohol abuse on transcriptomes of four brain cell classes. Direction and magnitude of alcohol-induced changes were estimated by plotting t distributions for genes enriched in a specific cell type (cell-type specificity was determined based on Cahoy et al., 2008; Oldham et al., 2008). A t value represents the magnitude and direction of alcohol effects ($t < 0$ indicates downregulation; $t > 0$ indicates upregulation by alcohol). Left column, Average t distributions for each cell class for three brain regions. Right column, Corresponding average \pm SEM t values (* $p < 0.05$; based on one-sample t test comparing average t values to zero chance with a Bonferroni's correction). Gray arrow points to a bump on cortical distribution caused by a cluster of alcohol upregulated genes (all $t > 2$).

$p < 0.006$). These proteins are partners in the NuRD (nucleosome remodeling and histone deacetylation) transcription corepressor complex (TCC) that is involved in transcriptional repression via coupling histone deacetylase (HDAC) activity with methylated DNA and establishing a repressive chromatin state (McDonel et al., 2009). Strikingly, most other members of the TCCs were also upregulated in alcoholics; these include *SIN3A*, *SIN3B*, *MTA1*, *MTA2*, *RBBP4*, *GATAD2A*, and *GATAD2B*, suggesting that these complexes are activated and play a role in downregulation of some genes in alcoholic brain.

Identification of candidate genes for alcohol addiction using a systems approach

Identification of candidate genes for human diseases remains the strategy of choice for genome-wide surveys, such as microarrays and genome-wide association studies. One goal of our systems approach was to establish a functional framework for prioritization of candidate genes. Gene coexpression network analysis de-

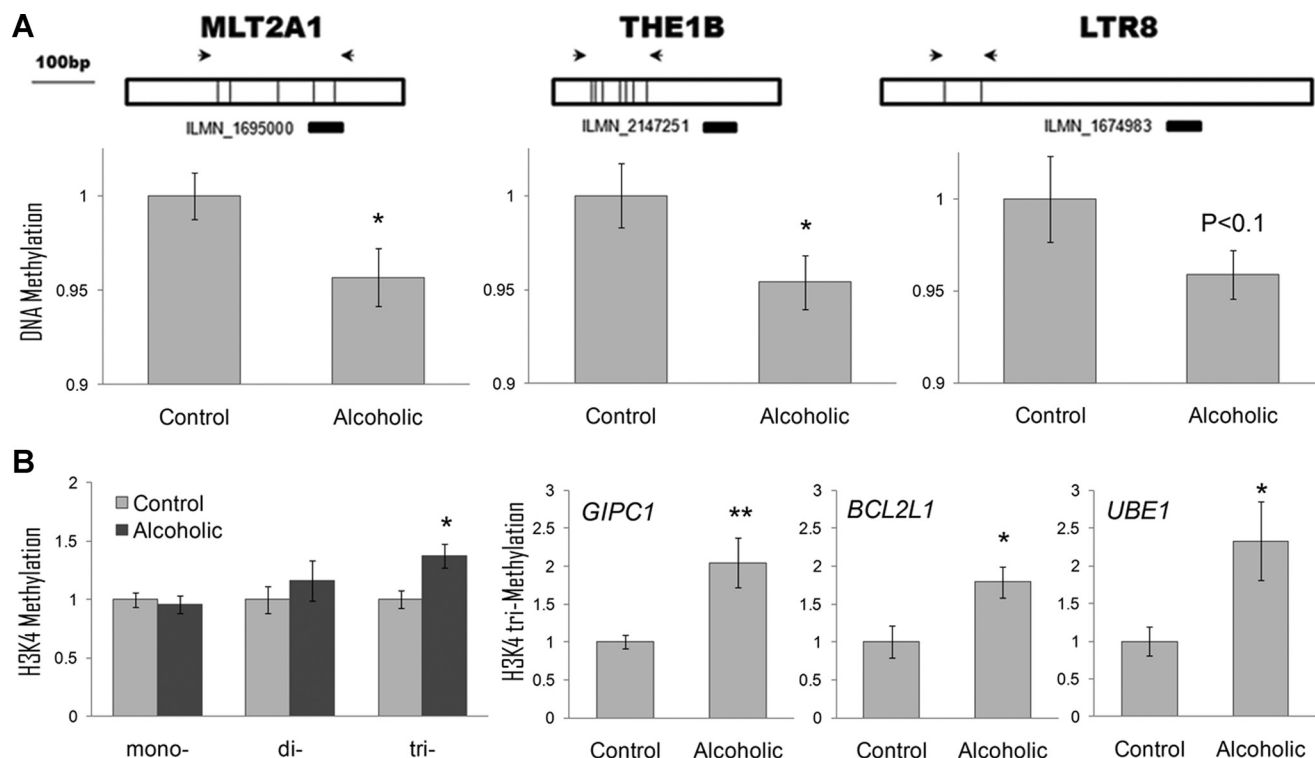


Figure 6. Epigenetic modifications in alcoholic cortex. **A**, DNA hypomethylation is associated with higher expression of LTR retrotransposons in alcoholics. Top, Schematic diagram of three families of LTR (MLT2A1, THE1B, and LTR8). Illumina probes shown below as black bars can detect expression of these classes of LTRs and were upregulated in alcoholic brain (fold change range from 20 to 48%; p value range from 0.007 to 0.03). Arrows show locations of RT-PCR primers used for DNA methylation assays. DNA methylation shown as fold change compared with control group was measured using a combination of DNA digestion with a methylation-sensitive enzyme and qRT-PCR. **B**, H3K4me3 shown as fold change compared with control group is increased in alcoholics. Left panel, Global methylation of H3K4 measured using a combination of ChIP and qRT-PCR. Three right panels, H3K4me3 levels at the promoter of three hub genes from the *ctx7* GC-rich module (gene symbols are italicized). * $p < 0.05$; ** $p < 0.01$, as determined by a t test.

termines a connectivity measure for individual genes based on their Pearson's correlations with all of the other genes in the module. This measure provides an estimate of the importance of the gene in gene networks, because highly connected hub genes proved to be functionally significant (Horvath et al., 2006). We nominated candidate genes based on two criteria: (1) differential expression between alcoholics and controls in, at least, one brain region and (2) being in the top 20% of hub genes with the highest intra-modular connectivity. We hypothesize that these genes have high functional significance in biological processes associated with alcohol addiction. Our finding that alcohol abuse changes gene expression through changes in chromatin states provided rationale for giving additional priority to genes involved in epigenetic regulation of gene expression.

For example, two histone methyltransferases, *MLL4* and *SETD1A*, were among alcohol-regulated hub genes in CTX and BLA, respectively, providing additional support for the importance of H3K4me3 in establishing patterns of gene expression in alcoholic brain. Another hub gene, *TRIM28* (KRAB-associated protein 1, KAP1) is 20–40% upregulated in alcoholics in all brain regions (all $p < 0.004$). The product of this gene is critical for silencing ERVs during early embryonic development (Rowe et al., 2010), and its upregulation may indicate the compensatory response of the cell to ERV activation. Methylation of DNA and other transmethylation reactions rely on the availability of SAM (S-adenosylmethionine) molecule, the primary methyl group donor in the cell. One of the hub genes downregulated in alcoholics in all brain regions was *MAT2B* (methionine adenosyltransferase II, β subunit; all $p < 0.004$), the enzyme involved in

the synthesis of SAM from methionine. The β subunit changes kinetic properties of the catalytic α subunit by rendering it more susceptible to product inhibition by SAM, and a downregulation of *MAT2B* in T cells was accompanied by a 6- to 10-fold increase in intracellular SAM levels (LeGros et al., 2001). Because SAM levels are decreased in alcoholics (Blasco et al., 2005), the downregulation of *MAT2B* in alcoholic brain may indicate a compensatory response to this reduction. In addition, several cortical genes acting at glutamatergic synapse, including *GRIN1*, *STX1A*, *SYN*, *DNM1*, *GRIK5*, *GRINA*, *VAMP2*, *GIPC1*, and *MIB2* were among the significantly upregulated hub genes (all $p < 0.05$), suggesting a central role of glutamate neurotransmission in alcohol dependence. Differential expression of several prioritized genes, including *DNMT1*, *MBD3*, *MLL4*, *SETD1A*, and *GIPC1* was further validated using qRT-PCR (Table 2). Overall, this analysis provides rationale for targeting epigenetic processes, glutamatergic synapse, and functionally relevant individual genes to promote the development of new therapies for human alcoholism.

Discussion

We used a novel systems approach to transcriptome profiling and provided the first comprehensive assessment of gene expression changes in alcoholic brain at a systems level. This approach allowed us to generate several systems hypotheses with an emphasis on epigenetic regulation of gene expression, and we obtained functional evidence for two of these hypotheses experimentally. Our results provide a functional framework for integrating data across alcohol-related studies, which we used to generate a global

Table 2. Results of qRT-PCR validation

Brain region	Gene symbol	Illumina ID	Taqman assay ID	Array <i>p</i> value	Array FC	Taqman <i>p</i> value	Taqman FC
BLA	<i>GIPC1</i>	ILMN_1796177	Hs00991802_m1	7.4E-05	1.65	4.9E-03	1.91
BLA	<i>SETD1A</i>	ILMN_1714327	Hs00986924_g1	4.1E-04	1.28	5.0E-02	1.27
BLA	<i>MLL4</i>	ILMN_1715968	Hs00998464_gH	2.7E-02	1.11	2.6E-02	1.26
BLA	<i>DNMT1</i>	ILMN_1760201	Hs02558036_s1	2.4E-03	−1.27	2.7E-03	−1.44
BLA	<i>MBD3</i>	ILMN_1698491	Hs00924265_g1	1.4E-06	1.42	3.8E-04	2.38
CTX	<i>GIPC1</i>	ILMN_1796177	Hs00991802_m1	4.9E-03	1.55	1.8E-02	1.93
CTX	<i>DNMT1</i>	ILMN_1760201	Hs02558036_s1	4.2E-02	−1.21	3.1E-02	−1.32

Array *p* values are based on a Bayesian two-tailed *t* test, and TaqMan *p* values are based on a Student's one-tailed *t* test. ID, Identification number; FC, fold change. Positive FC values indicate upregulation in alcoholics.

systems hypothesis for the role of chromatin modifications in alcohol dependence that consolidates the epigenetic regulation of gene expression and cellular changes in alcoholic brain (Fig. 7). We hypothesize that neuropathology and neuroadaptations that contribute to alcohol addiction and dependence are, at least in part, mediated by alcohol-induced epigenetically mediated changes in gene expression. Below we discuss the evidence for individual components of this hypothesis and the rationale for their integration.

Central to our hypothesis is the potentially critical role of DNA hypomethylation in alcohol-induced gene expression. We detected a small, but reliable, decrease in methylation of HERV sequences associated with a much larger increase in HERV transcript abundance, suggesting that a small decrease in DNA methylation can have profound effects on global gene expression. ERVs are heavily methylated in mammalian genomes, accounting for a large fraction of all methylated cytosines (Walsh et al., 1998), and an increase in ERV transcription is a sensitive marker of global DNA hypomethylation (Schulz et al., 2006; Balada et al., 2009). One example of the relationship between DNA methylation and ERV activity is that coat color in the yellow agouti mouse is controlled by the level of DNA methylation of LTR, upstream of the agouti locus (Morgan et al., 1999; Waterland and Jirtle, 2003). Methyl supplements, including extra folates, vitamin B₁₂, choline, and betaine, fed to dams increased the level of DNA methylation in the agouti LTR and changed the phenotype of offspring from yellow to mottled to pseudoagouti (Waterland and Jirtle, 2003). This research shows that subtle changes in DNA methylation are proportional to the level of activation of ERVs, which is consistent with our findings. Two other observations provide additional support for global DNA hypomethylation in alcoholic brain. First, a downregulation of *DNMT1* in alcoholic brain is consistent with literature showing similar response in some hypomethylating states associated with cancer and other pathological conditions (Hervouet et al., 2010). Second, upregulation of several ribosomal modules (Fig. 3) suggests a release of transcriptional repression of ribosomal DNA repeats by DNA hypomethylation (McStay and Grummt, 2008). Alcohol-induced global DNA hypomethylation has been reported in several

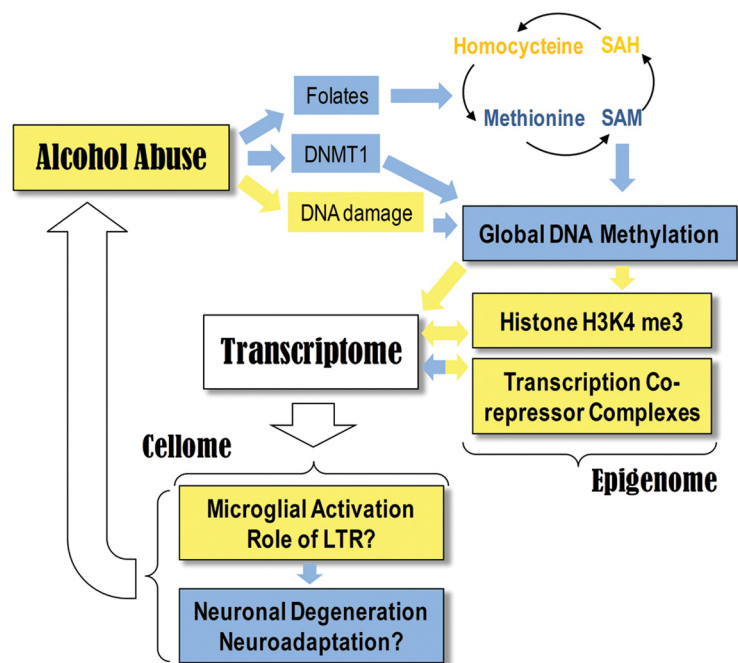


Figure 7. A systems hypothesis for the central role of epigenetic modifications in alcohol dependence. Yellow color indicates general increase, upregulation, or activation, whereas blue color indicates general decrease, downregulation, or degeneration. Chronic alcohol causes well-documented vitamin B and folate deficiencies that negatively affect one-carbon metabolism and can result in homocysteinemia and a decreased production of SAM, the methyl group donor in most transmethylation reactions. Decreased SAM and other alcohol-mediated effects, such as acetaldehyde-induced inhibition of the maintenance DNA methyltransferase DNMT1 and 5-methylcytosine demethylation induced by the DNA damage and repair can cause global DNA hypomethylation, a chromatin state associated with many pathological conditions, including cancer. DNA hypomethylation may trigger a chain of events resulting in changes in chromatin state, such as increase in H3K4 trimethylation and activation of TCCs, which result in changes in global gene expression. Epigenome-mediated changes in transcriptome can determine cell-type-specific functional states, such as activation of microglia, neuronal degeneration in the amygdala, and neuroadaptations in the PFC. In summary, alcohol-induced epigenetically mediated changes in gene expression may underlie brain pathology and brain plasticity associated with alcohol abuse and alcohol dependence.

peripheral tissues of alcohol-related animal models, in which it was proposed to play a role in alcoholic liver disease, fetal alcohol syndrome, and colon cancer (Garro et al., 1991; Choi et al., 1999; Lu et al., 2000; Shukla et al., 2008; Hamid et al., 2009). Although the effects of alcohol on DNA methylation and expression of individual genes in the CNS has been reported previously (Bleich and Hillemecher, 2009), our study is the first to demonstrate global changes in DNA methylation in alcoholic brain, in which it may contribute to the development and maintenance of alcohol dependence. Chronic alcohol can result in a decrease in DNA methylation via several mechanisms (Fig. 7), including vitamin B and folate deficiencies, resulting in an impairment of one carbon metabolism and a decrease in SAM (Hamid et al., 2009), acetaldehyde-mediated inhibition of the activity of DNMT1 (Garro et al., 1991), and 5-methylcytosine demethylation induced by alcohol-related DNA damage (Chen et al., 2011).

Specifically, SAM is decreased, whereas its metabolites S-adenosylhomocysteine and homocysteine are increased in chronic alcoholics (Blasco et al., 2005), which may be one cause of the global DNA hypomethylation.

Many chromatin modifications are mechanistically linked (Jaenisch and Bird, 2003), resulting in a limited number of chromatin states (Ernst et al., 2011). In addition to DNA hypomethylation, we detected an increase in global and gene-specific trimethylation of H3K4 and activation of several genes involved in TCCs in alcoholic brain, with all these chromatin modifications being associated with the methylation status of cytosines within CpGs. Both H3K4me3 and the HDAC activity coupled to TCCs can be changed by drugs of abuse (Pandey et al., 2008; Renthall and Nestler, 2009; Zhou et al., 2011). Specifically, acute ethanol and cocaine decrease HDAC activity, whereas ethanol withdrawal and chronic cocaine increase it (Pandey et al., 2008; Renthall and Nestler, 2009). Consistent with these findings and our own results is an upregulation of *MBD3* in brains of alcoholics and cocaine abusers (Liu et al., 2006; Zhou et al., 2011), which is a TCC protein critical for coupling HDAC activity and chromatin remodeling. Importantly, drug effects on key epigenetic “master regulators” result in changes in chromatin state that cause changes in global gene expression, some of which are molecular determinants of functional changes in brains of drug addicts. Drugs targeting these master switches are emerging as potential therapeutics for neurodegenerative disorders and drug addiction (Abel and Zukin, 2008; Renthall and Nestler, 2009).

Another important component of our hypothesis is the transcriptional activation of HERVs in alcoholic brain induced by DNA hypomethylation. Activation of ERVs has been linked to chronic diseases, including cancer, multiple sclerosis, and autoimmune disorders (Balada et al., 2009). It appears that this activation is not just a marker of global DNA hypomethylation but can result in functional consequences because an ERV-encoded glycoprotein, syncytin, can directly activate microglia and astrocytes and produce neuroinflammation (Antony et al., 2004). Microglial activation can result in neuronal degeneration (Crews et al., 2011), and compounds secreted by syncytin-activated astrocytes can produce cytotoxicity to oligodendrocytes and myelin degeneration (Antony et al., 2004), which is consistent with pathologies observed in alcoholics (Harper et al., 2003; Pfefferbaum et al., 2009; Zahr et al., 2011). Alcohol-induced neuroimmune response was recently proposed to be a critical factor in alcohol addiction (Crews et al., 2011), and we propose a potential role for ERVs in neuroinflammation and brain pathophysiology of human alcoholism.

Our global profiles of cell-type-specific genes were mainly consistent with microglial activation in all brain regions and neuronal degeneration in the amygdala. In addition, we detected a subset of synaptic genes highly upregulated in cortex of alcohol abusers. This upregulation may indicate an adaptive increase in glutamatergic transmission in response to the loss of neurons or alcohol-induced inhibition of NMDA receptors, which is consistent with literature showing general potentiation of glutamatergic synapses after chronic alcohol (Gass and Olive, 2008; Henriksson et al., 2008). The glutamatergic potentiation in the PFC may underlie the long-lasting impairment in cognitive control of goal-directed behaviors that characterizes addicted individuals (Koob and Volkow, 2010).

One unexpected finding of our study was that the nucleotide composition of a gene, measured as gene GC content, could determine, at least in part, its patterns of expression and regulation

by homeostatic perturbations, such as chronic alcohol. The importance of genome GC content in regulation of gene expression is well established because many transcription factors and other DNA-binding proteins that are part of chromatin modification complexes preferentially bind to GC- or AT-rich motifs (Lorincz et al., 2004; Zhang et al., 2004; Dekker, 2007; McDonel et al., 2009). It is possible that upregulation of many GC-rich genes and downregulation of many GC-poor genes in alcoholic brain were mediated by some of these DNA-binding regulators. How exactly nucleotide composition of a given gene contributes to its coexpression patterns and regulation by alcohol abuse will be addressed in future studies.

Our study was not designed to address causality in our integrative view of alcohol dependence. Therefore, alternative interpretations of our results are possible. For example, the observed chromatin changes may be secondary to the primary effects of chronic alcohol on different cells. The cause-and-effect relationships between different components of our systems hypothesis will be addressed by validation experiments in the future. Another limitation is that we cannot distinguish gene expression changes produced by chronic alcohol abuse and those associated with preexisting conditions, such as genetic polymorphisms or pathological states that may lead to alcoholism. Multiple studies in humans and animal models highlighted the importance of the genetic component in alcohol addiction (for review, see Crabbe, 2008; Mayfield et al., 2008; Spanagel, 2009; Treutlein and Rietschel, 2011). In an attempt to determine genes that may be regulated by genetic differences, we checked our top candidate genes against the Genetic Association Database (<http://geneticassociationdb.nih.gov/>), which is an archive of human genetic association studies of complex diseases, including brain pathologies and substance abuse syndromes. None of the chromatin modifying genes was genetically associated with alcoholism or other substance abuse diseases, and only the NMDA receptor NR1 subunit (*GRIN1*) was associated with alcoholism (Wernicke et al., 2003). Additional support for nongenetic causes of differential gene expression in our study is provided by a recent report showing that chronic intermittent alcohol consumption changes gene expression in brain of genetically homogeneous C57BL/6 mice (Wolstenholme et al., 2011). This study showed that chromatin modifications and glutamate signaling were top functional groups overrepresented with alcohol-related genes, which is consistent with our findings. Finally, a correlational analysis between alcohol variables and alcohol-related modules showed that the first eigengene of the *ctx12* LTR module (Fig. 3) upregulated in alcoholics was significantly correlated with the duration of drinking (corrected $p = 0.03$), suggesting that DNA hypomethylation and the upregulation of LTR retrotransposons is a consequence of chronic alcohol and not a preexisting condition. The combined evidence suggests that, in our study, global changes in gene expression in alcoholic brain are mainly caused by chronic alcohol abuse and that alcohol abuse changes gene expression via changes in chromatin states. To better understand the interplay between genetic, epigenetic, and environmental causes in controlling gene expression in alcoholism, integrative approaches across studies are warranted.

In summary, to our knowledge, our study is the first to present an integrated view of alcohol dependence using a systems approach to transcriptome profiling in human brain. The systems analysis of the transcriptome allowed us to make mechanistic predictions about the upstream epigenetic control as well as downstream cellular physiology. One implication is that epige-

netic interventions may effectively correct the widespread changes in brain gene expression and functional abnormalities produced by chronic alcohol abuse. Many epigenetic therapeutics have been developed for other diseases, and our study may direct some of these therapeutics toward alcoholism and drug addiction.

Notes

Supplemental material for this article is available at <http://www.utexas.edu/research/wcaar/manuscripts.html>. Supplemental Table 1 shows case and clinical information. Supplemental Table 2 shows overall results of WGCNA combined with the differential expression analysis. Supplemental Table 3 shows results of DAVID overrepresentation analysis in BLA. Supplemental Table 4 shows results of DAVID overrepresentation analysis in CNA. Supplemental Table 5 shows results of DAVID overrepresentation analysis in CTX. Supplemental Table 6 shows Illumina probes mapped to TEs. Supplemental Table 7 shows modular overrepresentation with cell classes and TEs. This material has not been peer reviewed.

References

- Abel T, Zukin RS (2008) Epigenetic targets of HDAC inhibition in neurodegenerative and psychiatric disorders. *Curr Opin Pharmacol* 8:57–64.
- Antony JM, van Marle G, Opii W, Butterfield DA, Mallet F, Yong VW, Wallace JL, Deacon RM, Warren K, Power C (2004) Human endogenous retrovirus glycoprotein-mediated induction of redox reactants causes oligodendrocyte death and demyelination. *Nat Neurosci* 7:1088–1095.
- Balada E, Ordi-Ros J, Vilardell-Tarrés M (2009) Molecular mechanisms mediated by human endogenous retroviruses (HERVs) in autoimmunity. *Rev Med Virol* 19:273–286.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 57:289–300.
- Blasco C, Caballeria J, Deulofeu R, Llagoña A, Parés A, Lluís JM, Gual A, Rodés J (2005) Prevalence and mechanisms of hyperhomocysteinemia in chronic alcoholics. *Alcohol Clin Exp Res* 29:1044–1048.
- Bleich S, Hillemecher T (2009) Homocysteine, alcoholism and its molecular networks. *Pharmacopsychiatry* 42 [Suppl 1]:S102–S109.
- Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, Xing Y, Lubischer JL, Krieg PA, Krupenko SA, Thompson WJ, Barres BA (2008) A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* 28:264–278.
- Chen CH, Pan CH, Chen CC, Huang MC (2011) Increased oxidative DNA damage in patients with alcohol dependence and its correlation with alcohol withdrawal severity. *Alcohol Clin Exp Res* 35:338–344.
- Choi SW, Stickel F, Baik HW, Kim YI, Seitz HK, Mason JB (1999) Chronic alcohol consumption induces genomic but not p53-specific DNA hypomethylation in rat colon. *J Nutr* 129:1945–1950.
- Crabbe JC (2008) Review. Neurogenetic studies of alcohol addiction. *Philos Trans R Soc Lond B Biol Sci* 363:3201–3211.
- Crews FT, Zou J, Qin L (2011) Induction of innate immune genes in brain create the neurobiology of addiction. *Brain Behav Immun* 25 [Suppl 1]:S4–S12.
- Day JJ, Sweatt JD (2012) Epigenetic treatments for cognitive impairments. *Neuropsychopharmacology* 37:247–260.
- Dekker J (2007) GC- and AT-rich chromatin domains differ in conformation and histone modification status and are differentially modulated by Rpd3p. *Genome Biol* 8:R116.
- Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, Bupp S, Shrestha P, Shah RD, Doughty ML, Gong S, Greengard P, Heintz N (2008) Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* 135:749–762.
- Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473:43–49.
- Garro AJ, McBeth DL, Lima V, Lieber CS (1991) Ethanol consumption inhibits fetal DNA methylation in mice: implications for the fetal alcohol syndrome. *Alcohol Clin Exp Res* 15:395–398.
- Gass JT, Olive MF (2008) Glutamatergic substrates of drug addiction and alcoholism. *Biochem Pharmacol* 75:218–265.
- Geschwind DH, Konopka G (2009) Neuroscience in the era of functional genomics and systems biology. *Nature* 461:908–915.
- Hamid A, Wani NA, Kaur J (2009) New perspectives on folate transport in relation to alcoholism-induced folate malabsorption: association with epigenome stability and cancer development. *FEBS J* 276:2175–2191.
- Harper C, Dixon G, Sheedy D, Garrick T (2003) Neuropathological alterations in alcoholic brains. Studies arising from the New South Wales Tissue Resource Centre. *Prog Neuropsychopharmacol Biol Psychiatry* 27: 951–961.
- Henriksson R, Kuzmin A, Okvist A, Harper C, Sheedy D, Garrick T, Yakovleva T, Bakalkin G (2008) Elevated synaptophysin I in the prefrontal cortex of human chronic alcoholics. *Synapse* 62:829–833.
- Hervouet E, Lallier L, Debieu E, Cheray M, Geairon A, Rogniaux H, Lousouarn D, Martin SA, Vallette FM, Cartron PF (2010) Disruption of Dnmt1/PCNA/UHRF1 interactions promotes tumorigenesis from human and mice glial cells. *PLoS One* 5:e11333.
- Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Feliciano RM, Laurance MF, Zhao W, Qi S, Chen Z, Lee Y, Scheck AC, Liao LM, Wu H, Geschwind DH, Febo PG, Kornblum HI, Cloughesy TF, Nelson SF, Mischel PS (2006) Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci U S A* 103:17402–17407.
- Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* [Suppl] 33:245–254.
- Jurd R, Thornton C, Wang J, Luong K, Phamluong K, Kharazia V, Gibb SL, Ron D (2008) Mind bomb-2 is an E3 ligase that ubiquitinates the N-methyl-D-aspartate receptor NR2B subunit in a phosphorylation-dependent manner. *J Biol Chem* 283:301–310.
- Koch CM, Andrews RM, Flicek P, Dillon SC, Karaöz U, Clelland GK, Wilcox S, Beare DM, Fowler JC, Couttet P, James KD, Lefebvre GC, Bruce AW, Dovey OM, Ellis PD, Dhami P, Langford CF, Weng Z, Birney E, Carter NP, Vetric D, Dunham I (2007) The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res* 17:691–707.
- Koob GF, Volkow ND (2010) Neurocircuitry of addiction. *Neuropsychopharmacology* 35:217–238.
- LeGros L, Halim AB, Chamberlin ME, Geller A, Kotb M (2001) Regulation of the human MAT2B gene encoding the regulatory beta subunit of methionine adenosyltransferase, MAT II. *J Biol Chem* 276:24918–24924.
- Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen TM, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, Desaki AL, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445:168–176.
- Liu J, Lewohl JM, Harris RA, Iyer VR, Dodd PR, Randall PK, Mayfield RD (2006) Patterns of gene expression in the frontal cortex discriminate alcoholic from nonalcoholic individuals. *Neuropsychopharmacology* 31:1574–1582.
- Lorincz MC, Dickerson DR, Schmitt M, Groudine M (2004) Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat Struct Mol Biol* 11:1068–1075.
- Lu SC, Huang ZZ, Yang H, Mato JM, Avila MA, Tsukamoto H (2000) Changes in methionine adenosyltransferase and S-adenosylmethionine homeostasis in alcoholic rat liver. *Am J Physiol Gastrointest Liver Physiol* 279:G178–G185.
- Mayfield RD, Harris RA, Schuckit MA (2008) Genetic factors influencing alcohol dependence. *Br J Pharmacol* 154:275–287.
- Maze I, Feng J, Wilkinson MB, Sun H, Shen L, Nestler EJ (2011) Cocaine dynamically regulates heterochromatin and repetitive element unsilencing in nucleus accumbens. *Proc Natl Acad Sci U S A* 108:3035–3040.
- McDonel P, Costello I, Hendrich B (2009) Keeping things quiet: roles of NuRD and Sin3 co-repressor complexes during mammalian development. *Int J Biochem Cell Biol* 41:108–116.
- McStay B, Grummt I (2008) The epigenetics of rRNA genes: from molecular to chromosome biology. *Annu Rev Cell Dev Biol* 24:131–157.
- Miller JA, Oldham MC, Geschwind DH (2008) A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J Neurosci* 28:1410–1420.
- Miller JA, Horvath S, Geschwind DH (2010) Divergence of human and

- mouse brain transcriptome highlights Alzheimer disease pathways. *Proc Natl Acad Sci U S A* 107:12698–12703.
- Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* 23:314–318.
- Mulligan MK, Rhodes JS, Crabbe JC, Mayfield RD, Adron Harris R, Ponomarev I (2011) Molecular profiles of drinking alcohol to intoxication in C57BL/6J mice. *Alcohol Clin Exp Res* 35:659–670.
- Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH (2008) Functional organization of the transcriptome in human brain. *Nat Neurosci* 11:1271–1282.
- Pandey SC, Ugale R, Zhang H, Tang L, Prakash A (2008) Brain chromatin remodeling: a novel mechanism of alcoholism. *J Neurosci* 28:3729–3737.
- Pfefferbaum A, Rosenbloom M, Rohlfing T, Sullivan EV (2009) Degradation of association and projection white matter systems in alcoholism detected with quantitative fiber tracking. *Biol Psychiatry* 65:680–690.
- Ponomarev I, Rau V, Eger EI, Harris RA, Fanselow MS (2010) Amygdala transcriptome and cellular mechanisms underlying stress-enhanced fear learning in a rat model of posttraumatic stress disorder. *Neuropsychopharmacology* 35:1402–1411.
- Renthal W, Nestler EJ (2009) Chromatin regulation in drug addiction and depression. *Dialogues Clin Neurosci* 11:257–268.
- Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Mailard PV, Layard-Liesching H, Verp S, Marquis J, Spitz F, Constam DB, Trono D (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* 463:237–240.
- Saris CG, Horvath S, van Vught PW, van Es MA, Blauw HM, Fuller TF, Langfelder P, DeYoung J, Wokke JH, Veldink JH, van den Berg LH, Ophoff RA (2009) Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. *BMC Genomics* 10:405.
- Schulz WA, Steinhoff C, Florl AR (2006) Methylation of endogenous human retroelements in health and disease. *Curr Top Microbiol Immunol* 310:211–250.
- Shukla SD, Velazquez J, French SW, Lu SC, Ticku MK, Zakhari S (2008) Emerging role of epigenetics in the actions of alcohol. *Alcohol Clin Exp Res* 32:1525–1534.
- Slotkin RK, Martienssen R (2007) Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* 8:272–285.
- Spanagel R (2009) Alcoholism: a systems approach from molecular physiology to addictive behavior. *Physiol Rev* 89:649–705.
- Sugino K, Hempel CM, Miller MN, Hattox AM, Shapiro P, Wu C, Huang ZJ, Nelson SB (2006) Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nat Neurosci* 9:99–107.
- Torkamani A, Dean B, Schork NJ, Thomas EA (2010) Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. *Genome Res* 20:403–412.
- Treutlein J, Rietschel M (2011) Genome-wide association studies of alcohol dependence and substance use disorders. *Curr Psychiatry Rep* 13:147–155.
- Vinogradov AE (2005) Dualism of gene GC content and CpG pattern in regard to expression in the human genome: magnitude versus breadth. *Trends Genet* 21:639–643.
- Walsh CP, Chaillet JR, Bestor TH (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 20:116–117.
- Waterland RA, Jirtle RL (2003) Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol* 23:5293–5300.
- Wernicke C, Samochowiec J, Schmidt LG, Winterer G, Smolka M, Kucharska-Mazur J, Horodnicki J, Gallinat J, Rommelspacher H (2003) Polymorphisms in the *N*-methyl-D-aspartate receptor 1 and 2B subunits are associated with alcoholism-related traits. *Biol Psychiatry* 54:922–928.
- Wolstenholme JT, Warner JA, Capparuccini MI, Archer KJ, Shelton KL, Miles MF (2011) Genomic analysis of individual differences in ethanol drinking: evidence for non-genetic factors in C57BL/6 mice. *PLoS One* 6:e21100.
- Yi Z, Petralia RS, Fu Z, Swanwick CC, Wang YX, Prybylowski K, Sans N, Vicini S, Wenthold RJ (2007) The role of the PDZ protein GIPC in regulating NMDA receptor trafficking. *J Neurosci* 27:11663–11675.
- Zahr NM, Kaufman KL, Harper CG (2011) Clinical and pathological features of alcohol-related brain damage. *Nat Rev Neurol* 7:284–294.
- Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4:Article17.
- Zhang L, Kasif S, Cantor CR, Broude NE (2004) GC/AT-content spikes as genomic punctuation marks. *Proc Natl Acad Sci U S A* 101:16855–16860.
- Zhou Z, Yuan Q, Mash DC, Goldman D (2011) Substance-specific and shared transcription and epigenetic changes in the human hippocampus chronically exposed to cocaine and alcohol. *Proc Natl Acad Sci U S A* 108:6626–6631.